

Assessments of the Dangerousness of Mental Patients Held in Maximum Security

Vernon L. Quinsey*

Problems in the identification of men who are likely to commit serious aggressive crimes against persons have inspired a large number of recent reviews (Gulevich & Bourne, 1970; Megargee, 1976; Mesnikoff & Lauterback, 1975; Monahan, 1975; Quinsey, 1977a, 1977b; Quinsey, Ambtman, & Pruesse, 1977; Shah, 1978; Steadman, 1976; Steadman & Cocozza, 1974; Wenk, Robinson, & Smith, 1972). The quantity of reviews attests both to the importance of this unresolved problem and to the lack of need for yet another review of the literature on the prediction of dangerousness.

The purpose of the present paper is not to present a survey of the literature but to describe the progress of a research program on dangerousness which has been conducted at the all male, maximum security, "Oak Ridge" Division of the Mental Health Centre in Penetanguishene, Ontario since 1971. A series of empirical studies involving assessments of the dangerousness of mental patients housed in maximum security will be reviewed from clinical, demographic, behavioral, psychometric, and psychophysiological perspectives in turn. It is hoped that this description of a series of inter-related research projects will indicate where progress has, and has not, been made and, in so doing, point to directions for future research. Because Oak Ridge is a psychiatric institution, many of the research studies deal with various measures of "mental illness" on the assumption (Quinsey, 1977a) that the psychiatric problems of Oak Ridge patients are related to their violent or antisocial behaviors. Similarly, because certain of the more retarded and/or psychotic patients are frequently assaultive within the institution much of our research has dealt with intra-institutional dangerousness.

Clinical Assessments

The interdisciplinary conference model of clinical assessment is the most widely used in deciding who should be released from maximum security psychiatric institutions. This model has the advantages of diffusing responsibility for decision making to a limited degree and of providing, under ideal circumstances, for the synthesis of observations from a variety of perspectives. The outcomes of such conferences are of great importance in psychiatric institu-

*Director of Research, Mental Health Centre, Penetanguishene, Ontario L0K 1P0, Canada. Portions of the research reported in this paper have been funded by the Ontario Mental Health Foundation, the Federal Department of the Solicitor General, and the Laidlaw Foundation. I wish to thank M. Pruesse, M. Rice, and G. Varney for their comments on an earlier draft of this paper.

tions because most of the patients who are assessed are being detained on fully indeterminate bases.

There are a number of serious methodological problems in assessing whether these conferences result in accurate decisions as to the dangerousness of involuntarily detained mental patients. Chief among these is that all patients whom the conference thinks are dangerous are kept and only those perceived as not dangerous are released. Unfortunately, only those patients who are released can be followed up. Inferences about decision accuracy from follow-up studies, however, are difficult unless we assume that the patients who are released are similar to those who are retained. It would be difficult to defend the proposition that clinical conferences randomly select patients for release. It should be noted in this connection that the Baxstrom study (Steadman & Cocozza, 1974) did study "randomly" released patients but the population was very old at the time of discharge. Similar problems of inference occur in follow-up studies of offenders who have been on determinate sentences as well (Quinsey, Ambtman, & Pruesse, 1977), because inmates at the end of their sentences are different than those at the beginning and this difference is related to the amount of discretion exercised by parole boards and sentence length.

Because of the difficulty in obtaining a randomly released sample of patients, the accuracy of clinical conference decision making is more frequently assessed indirectly. Crucial to the logic of these indirect studies is the notion that, if assessors do not agree in their judgments, they can't all be accurate. In other words, inter-judge disagreement sets an upper limit to the accuracy (or the validity coefficient) that can be achieved. It can be argued, of course, that some clinicians may be extremely accurate and, hence, disagree with those clinicians who are less accurate and that the use of inter-clinician reliability indices thus obscure the fact that some clinicians really can predict dangerousness. From a practical viewpoint, however, this objection is beside the point because, if true, conference assessments remain inherently subjective and their outcome critically depends on the composition of the conference team. Until such time as we can identify "super clinicians" and use only their judgments, low inter-judge agreements imply low accuracy of decision-making and a subjective decision-making process. It should be clear that, although this discussion has focused on individual clinicians, the same argument holds for group decisions as well because the joint decision would depend upon who formed the group.

In addition to providing data on inter-judge reliability, conference studies can also be used to identify variables which the clinicians perceive as relevant to assessments of dangerousness. These variables can then be used in validational follow-up studies, which, it must be admitted, are subject to the methodological problems inherent in a selective release policy.

We have completed three studies of clinical conferences; a study of 39 pre-trial assessment conferences, a study of 105 conferences of men found not guilty by reason of insanity or unfit to stand trial, and an assessment study of 30 patients under artificially controlled conditions. In the remand conference study (Quinsey, 1975), it was found using questionnaires completed by each conference participant, that psychiatric attendants perceived the remands as significantly less dangerous and less likely to benefit from treatment than either physicians or other professional staff perceived them. Perceived dangerousness

was positively correlated with rated degree of mental illness, treatability within a maximum security psychiatric unit, and poor likelihood of remaining out of prisons or mental hospitals if released. Remands who had been charged with an offense against persons were perceived as more dangerous than those not so charged.

The significant occupational differences found in the remand study hinted at differences amongst individual judges which could not be pursued in that study because of its methodology. A larger conference study, involving patients found not guilty by reason of insanity or (infrequently) patients found unfit for trial was designed to assess the amount of inter-clinician congruence more directly (Quinsey & Ambtman, 1979a). In this study, three forensic psychiatrists and a psychologist filled out questionnaires (during or after the conference) regarding each patient to determine his eligibility for release. Inter-clinician correlations were calculated for each of ten rated patient variables and a stepwise multiple regression equation was calculated for each clinician, relating ten patient background variables taken from the files to his dangerousness ratings. It was found that patients were more likely to receive release recommendations if they had shown unambiguous premeditation of their offense, had received four or fewer progress notes in the preceding four month period which mentioned disciplinary problems or deterioration in their psychiatric condition, and if they were not receiving psychotropic medication at the time of conference. The dangerousness ratings of the four clinicians were, as expected, each highly related to the conference recommendation but the average inter-clinician correlation on the dangerousness ratings was .60, indicating a rather modest amount of agreement. The ten background variables were significantly correlated with each clinician's dangerousness ratings (yielding an average R of .48). In addition to the three background variables already mentioned the other seven variables were: number of admissions to corrections, a diagnosis of personality disorder (including sexual deviation), rated offense severity, number of previous admissions to Oak Ridge, months in Oak Ridge, age at the time of conference, and number of admissions to other psychiatric hospitals. After the study, three of the clinicians were asked to rank order the importance of these ten variables in arriving at an appraisal of a patient's dangerousness and to indicate the direction of the relationships; their rankings disagreed with each others' in terms of both the importance and direction of the variables' relationship to dangerousness. The ranking data, although gathered in a contrived manner, indicate that the clinicians did not have similar weighting strategies for combining the information to be used in assessment.

There are two methodological limitations in the conference study described above: the first is that the clinicians discussed the case prior to making their ratings, which presumably led to overestimates of inter-rater agreement, and the second is that the background variables selected from the files may not be those most highly related to dangerousness ratings. The final study of this series (Quinsey & Ambtman, 1979b) used an artificial assessment situation, instead of actual conferences, to examine (a) inter-clinician congruence without the benefit of previous discussion and (b) the contribution of various types of information to the assessment of patient dangerousness. Thirty patients were selected according to whether their admission offense fell into one of the follow-

ing three categories: a nonsexual assaultive offense against an adult (usually a murder), a sexual offense involving bodily contact with a child 13 years of age or younger, or an offense against property. Three types of information were gathered for each patient: offense description, previous history, and clinical assessment (including psychological testing, mental status, and progress in treatment). The information types overlapped only in that they all contained the patient's age and months of current institutionalization. The data were rated twice by each rater: on one occasion they received each type of information separately, as if they came from different patients, and, on another, the three types were presented together as one file (which they actually were). The two rating occasions were separated by a minimum of five weeks and the order of presentation varied over raters. Four forensic psychiatrists and nine high school teachers independently rated each of the three information types separately and together on the likelihood of an assaultive offense, likelihood of a property offense, and seriousness of an assaultive offense should one be committed. All judgments were made as if the patient were to be released at the time of assessment. In addition, raters indicated whether the patient should be released at the time of assessment.

Neither teachers nor psychiatrists showed high levels of inter-rater correlations on the three information types, whether presented separately or together, regardless of the rating they were asked to make. Despite the low inter-rater correlations, the correlations between the averages of the occupational groups tended to be quite high. The psychiatrists and teachers also did not agree among themselves as to which patients should be released. Regression equations were computed to predict the ratings of the total file from the ratings of the three information types (when presented separately) using first, the average of the psychiatrists and, second, the average of the teachers. It was found that the assessment data did not contribute to the ratings of the total file on any of the three dimensions and that only the offense description contributed to the rating of the total file on the dimension of "seriousness of an assaultive offense."

Taken together, these three studies indicate that clinical conferences cannot accurately predict patient dangerousness, that the clinical assessment data are not weighted much in arriving at such an assessment and that forensic psychiatrists probably make judgments very similar to those that would be made by any educated layperson.

Predictions from Demographic Data

In the preceding section we have seen that accurate decisions about which patients should be released are not likely to be made at clinical conferences. This inaccuracy could result, however, not because the data on which the decisions were based were invalid predictors in themselves but rather because the clinicians combined and weighted the data idiosyncratically (for a discussion of clinical versus actuarial prediction see Wiggins, 1973). It is, therefore, of interest to determine to what extent post-release behavior can be predicted from various sorts of data which are considered by the conference team. In this section, standard demographic and clinical data, which are always available to clinical conference teams, will be considered.

We have performed four follow-up studies of released Oak Ridge patients in an attempt to develop a prediction method using demographic variables. In the first of these (Quinsey, Warneford, Pruess, & Link, 1975), 92 civilly committed patients, who had been released from Oak Ridge by a review board after having been refused discharge by the hospital, were followed up for a one to four year period. Sixteen percent of these patients committed a post-release violent act against persons (which included threatening, assault, robbery with violence but not simple robbery or possession of a weapon) and a total of 38% were convicted of a new offense, readmitted to Oak Ridge, or both. It was found that patients who had committed a previous violent crime were more likely to commit a subsequent violent offense than other patients. Patients who were diagnosed as personality disordered were more likely to be returned to Oak Ridge or be convicted of a new offense than those diagnosed as psychotic.

In a subsequent study (Quinsey, Pruess, & Fernley, 1975a) we followed up 56 patients who had been treated and released after having been found by the courts to be not guilty by reason of insanity or unfit for trial. The patients in this study resembled those held for the same reasons in other areas of Canada (Quinsey & Boyd, 1977). The average follow-up period was 30 months. In sharp contrast to the patients in the above study, these patients had usually committed very serious crimes against persons and were almost always transferred to minimum security psychiatric institutions instead of the community. During the follow-up period, five of these patients were convicted of a new offense or were returned to Oak Ridge and only one committed a violent offense against persons. Needless to say, the low rate of violent recidivism obviated any attempt to identify predictors of violent behavior in this sample.

Using a broader cross section of released Oak Ridge patients, we devised a simple numerical score to predict failure, defined as a readmission to Oak Ridge or a conviction for a new offense (Quinsey, Pruess, & Fernley, 1975b). In this study, 20 civilly committed patients who were discharged by the hospital, 20 civilly committed patients who were released by the external board of review, and 20 patients who had been committed by the courts were followed up for an average of 39 months. We found that one third of the sample failed but very few committed violent offenses. A score was calculated for each patient by awarding him one point for each of the following five variables: diagnosis of personality disorder, under 31 years of age at time of release, having spent less than 5 years in psychiatric hospitals, not being sent to Oak Ridge for a violent offense, and not having lived until age 16 with both parents. Patients with scores of three or over were significantly more likely to fail. Again, the rate of violent recidivism was too low to permit its separate study.

In order to validate our prediction scale and to gather enough data to address the issue of violent recidivism, we studied all patients who had been treated in Oak Ridge and who were released in 1972 (Pruess & Quinsey, 1977). There were 206 men who met these criteria and they were followed up for a 37-49 month period. Forty-six percent of the sample failed (as defined above) and 17% of the total committed at least one violent offense against persons. The scale designed to predict failure correctly classified 65% of the sample (down from 78% in the original study). Patients who were under 31 at the time of discharge and who were diagnosed as personality disordered were both more likely

to fail and more likely to commit post-release violent offenses against persons.

In agreement with other research (Quinsey, Ambtman, & Pruesse, 1977), this series of follow-up studies indicates that the accurate prediction of which patients will fail using demographic variables is not feasible at the present time; it is also apparent that the prediction of violent recidivism is even more problematic because of the low base rate of the phenomenon. These results should be no surprise, however, as it would be highly naive to suppose that demographic information could be highly related to post-release offending because most persons identified by any combination of standard demographic and clinical variables do not commit violent offenses, thus the low base rate of violence leads inevitably to its overprediction. This conclusion, however, in no way means that follow-up research of this type is fruitless. Demographic data can be used to identify subgroups of patients for whom the rate of violent recidivism is high enough to make further predictive research possible. There is little to be gained by conducting predictive studies on subgroups of patients for whom the best prediction is that none of them will be violent because so few of them commit future violent offenses.

Lest we close this section on too optimistic a note to be fashionable among researchers in the area of dangerousness, however, it should be added that, to use the results of follow-up studies in a practical manner, it must be assumed that the patients who are released are representative of those who remain. The results of our first and third follow-up studies indicate that civilly committed patients who are released against the advice of the hospital have post-release records which are similar to those that the hospital itself released; this finding means that the assumption of similarity between released and retained patients is at least partially true but does not speak to the issue of the dangerousness of the patients that nobody thought should be released.

Psychometric Assessments

Data from psychological testing is routinely gathered on inmates newly admitted to correctional facilities and offenders admitted to psychiatric facilities. In the latter case, these data are used together with a mental status examination by the psychiatrists as well as other information in assessing the offender's potential dangerousness and treatability. The MMPI appears to be the instrument most commonly used because of its computer scoring capability and the existence of a great deal of normative data.

None of the original clinical scales of the MMPI have been specifically designed for use in the prediction of dangerousness, although among mentally ill offender populations it would be reasonable to assume some correspondence between clinical pathology as measured by the MMPI and antisocial behavior. In addition, new scales have been developed using MMPI items which have been related to antisocial behavior. Perhaps most interesting among these is the Overcontrolled-Hostility (O-H) scale developed by Megargee, Cook and Mendelsohn (1967). These investigators showed that the O-H scale could differentiate men who had committed an isolated murderous offense from those who had committed more numerous but less severe assaultive offenses.

We have replicated Megargee's essential finding (Arnold, Quinsey & Velner,

1977) by showing that admission O-H scores were higher among men found not guilty by reason of insanity, and housed in maximum security than a group composed of men housed in a minimum security psychiatric hospital who had been involuntarily held for psychiatric treatment, referred by the courts for assessment, or who had voluntarily committed themselves. This finding agrees with Megargee's because the men found not guilty by reason of insanity had usually committed very serious crimes against persons and usually did not have previous criminal histories. It was of interest that none of the maximum security patients who scored above their group's median O-H score had ever been admitted previously to a correctional facility.

These findings encouraged us to use the MMPI in a study of men remanded for a psychiatric examination who had been charged with murder or attempted murder of a family member or girlfriend, murder or attempted murder of a non-family member, arson, rape, child molesting, or a property offense (Quinsey & Arnold, 1978). There were 25 subjects per group. A multiple stepwise discriminant analysis was computed to predict group assignment from the standard MMPI variables, the O-H scale, and several demographic and clinical variables. The analysis was then repeated with the exception that the two murder and attempt murder groups were combined and then divided according to whether they had had a previous admission to corrections.

It was hypothesized that the murder family and arson groups would have the highest O-H scores but this hypothesis was not supported. Similarly, contrary to expectations, the murder subjects who had no previous admissions to corrections did not have higher O-H scores than the other subjects.

The most important variables in distinguishing amongst the groups in which murderers were divided as to family and nonfamily victims were "age on admission" and "whether in corrections before." When the murderers were categorized according to whether they had been in corrections before, the most discriminating variable was "whether diagnosed as personality disordered or not." The results of this study, therefore, support the follow-up studies in identifying age and diagnosis as important variables. The MMPI variables, as might be expected from the low weight given to assessment data in the conference studies, were relatively unimportant. It is not clear why the O-H scale was not related to group assignment.

Laboratory Operant Studies of Assaultive Men

Because aggressive behaviors are operants, it would be expected that they show similarities to other operants. More specifically, if we assume that frequent physical assaultiveness is related to some sort of inhibitory deficit on the part of the assailter, then highly assaultive patients might have difficulty with any operant task which requires suppression for efficient responding. There are both empirical and theoretical reasons to believe that frequently assaultive men do have problems with response suppression (Quinsey, Varney, & McCann, 1978). If such an inhibitory deficit could be measured using a laboratory task, we would be in an excellent position to study methods of reducing assaultiveness indirectly using precise operant laboratory methods.

In order to examine this approach, we selected 16 Oak Ridge patients who

had been the aggressor in at least 4 intra-institutional assaults in the 14 months prior to the study and compared them with 39 patients who had committed one or no assaults in the same period. Following pretraining on a concurrent schedule in which no reinforcement was available on one lever (extinction) and points were awarded on a fixed interval 60 sec schedule on the other, subjects were randomly assigned to one of three conditions. In the first condition (omission training) subjects were rewarded for not responding on the lever on which they had previously received reinforcement. In the second condition (reinforced alternative), subjects were rewarded on a fixed interval 60 sec schedule for responding on the previously nonreinforced lever. In the third condition (differential reinforcement of low rates), subjects were rewarded for spaced responding on the previously reinforced lever. No difference was found in response rate on any of the three response reduction schedules between assaultive and nonassaultive subjects. The two schedules which rewarded alternate behaviors (omission training and reinforced alternative) reduced response rate on the previously reinforced lever more than the differential reinforcement of low rates schedule. Subjects who made more qualitative errors on the Porteus Maze test (a measure of impulsiveness) showed less reduction in response rate than those subjects who made fewer such errors.

This operant study, as well as two previous similar studies which we conducted, failed to show inhibitory deficits among highly assaultive Oak Ridge patients. It does not appear, therefore, that assaultive patients have inhibitory deficits that are general with respect to the operant behavior studied. If laboratory research is to uncover such differences, it may be more profitable to examine social behaviors and responses that are more closely related to physical aggression and anger.

Ward Behavior

Intra-institutional behavior has sometimes been found to be related to post-release dangerousness and sometimes not; often the relationship exists but is not straightforward (Tong & McKay, 1959; Waller, 1974). The degree of the relationship is probably affected by the quality of the intra-institutional observations as well as the similarity of the institutional and post-release environments.

Although we have found that the number of negative progress notes in a patient's file is related to whether he is recommended for release or not, traditionally kept notes on patient progress are poor descriptions of a patient's real behavior. In an early study (Quinsey, 1972), the 11 descriptive phrases most commonly used in ward progress reports by attendant staff were identified: quiet and cooperative, demanding, good worker on the ward, very unpredictable, disturbed, hostile and threatening, very confused, noisy, manipulative, no management problem, and surly. Two attendants from each of four maximum security wards were asked to circle which of any of these descriptors applied to each of the patients (average $n = 34.25$) on their ward in the last 8 hours. The percent agreements were very high for the 44 comparisons. However, many of the items were seldom circled as applying to any patient. This occurred presum-

ably because most of the items referred to undesirable qualities or behaviors and most patients were well behaved and appeared relatively normal. The preponderance of negative items in the ward books resulted from the fact that most patients were not mentioned in the ward book unless they were misbehaving. To test the idea that many of the high agreements resulted from the attendants simply noting that the term was inapplicable to nearly all their patients, the percent agreement was calculated for each item using only those patients for whom the item had been endorsed by at least one of the attendant pair. Essentially the question being asked was: if an attendant describes a patient using a particular phrase, how likely is another attendant to agree with him? It was found that the percent agreements calculated in this manner were rather low with the exception of the three positive items. Although these data should be accepted with some caution due to the shrinkage of *ns* for the negative items, it does appear as though substantial disagreement existed among ward staff as to the applicability of the terms used in their ward books to particular patients.

The consequence of subjectivity in ward observations and the fact that patients are not usually mentioned in the ward books unless they are misbehaving is inevitably a gloomy caricature of the patient's behavior. A further problem with traditional ward observations is that they are noncomparable; for example, if a patient is described as "surly when examined by the physician" on one occasion and as a "good worker on the ward on another," we have no idea whether he has improved or not because the descriptors are relevant neither to the same behavior nor the same situation.

In order to gather more interpretable data based on ward behavior, we have conducted studies of patients involved in both a patient led milieu therapy program and a modified token economy system. In both types of programs we have attempted to obtain data that were objective, quantifiable, related at least in a *prima facie* manner to the patients' dangerousness, and sensitive to treatment effects.

In the milieu therapy study (Quinsey & Harris, 1976) we studied goal attainment ratings on one ward of the Social Therapy Unit (STU) of Oak Ridge, various aspects of which have been described elsewhere (Barker & Buck, 1977; Barker & Mason, 1968a, 1968b; Barker, Mason, & Wilson, 1969). Twenty-two patients were studied for a 5-month period. During this time their program was largely self contained and patient led. The program included long term interactions between pairs of patients, government by patient committees, the administration of drugs such as LSD and scopolamine in a therapeutic context to patient volunteers, and marathon small group interactions in an environment isolated from the rest of the ward.

All patients were male. Their mean age was 22.41 yrs ($SD = 5.65$) and the majority were diagnosed as personality or character disordered. The offense leading to their admission (for which they were not necessarily charged) was against persons in 82% of the cases, 11 patients had been charged with murder or manslaughter.

Extensive discussions were held with the STU staff to determine what variables they took into consideration when assessing patient change. After these

discussions, a group consensus was reached as to what goals were appropriate for each of the 22 patients. Practice ratings of "typical" patients indicated fair inter-rater agreement. The items on this scale in various combinations were measures of the dimensions on which the patients were expected to improve. Examples of these dimensions are: assertive-unassertive, irresponsible-responsible, confiding-withholding, and likeable-unlikeable, each on six point scales. There were 14 of these bipolar dimensions plus three other items (e.g., the amount of paranoid suspicion shown by a patient). All of the items of the scale were not relevant to each of the patients but each patient was rated on all of the items. With the exception of the ward supervisors and off-ward or professional staff, none of the raters knew which items applied to which patients; it is unlikely that even the staff members who selected the treatment goals knew in detail which items applied to which patients. The average number of goals set for the patients was 6.18 ($SD = 1.92$).

Ratings took place at the end of each month and covered the entire month's behavior, as staff felt that a lengthy period was required to obtain enough observational material for rating. Each of the 22 patients rated every other patient and himself. Three off-ward staff (two nurses and the chaplain) the psychiatrist unit director, and 10 on-ward attendant staff also rated each of the patients. Raters and patients were included only if they made ratings on each of the five months.

In order to determine inter-rater agreement, ratings were averaged within the following groups: on-ward staff, off-ward staff, and patients (excluding self ratings). When ratings were averaged within occupational groups and across patients, only goals that were relevant to all raters were included — i.e., "like others more" would be rated only by the patient himself. The patient's self ratings were not included in the patient ratings. The means of these groups and the self ratings and unit director ratings were inter-correlated using one randomly chosen goal for each patient. There was a moderate amount of inter-rater agreement and an increase in the amount of agreement between the first and fifth month. But, because averaging within groups artificially inflated the correlations, more conventional inter-rater reliabilities were also calculated. Two attendants and two patients were sampled randomly and two professional staff were chosen arbitrarily for this purpose. One of each patient's goals was randomly selected and a correlation was calculated between each pair's ratings of these randomly selected goals for both the first and fifth month. These correlations indicated modest to no agreement.

Unfortunately, even the modest agreement found within these ratings reflected a disturbing aspect of the data. The patients were typically rated high on all the items (scored in a favorable direction) both before and after their participation in the program. That is, inter-rater reliabilities partially reflected the tendency of the raters to use the high end of the scale. The high rating of the patients at the beginning of the program has two implications: patients couldn't show much improvement because they were near the top of the scale initially, and doubt was cast on the selection of the goals.

In order to examine whether any change occurred, all of each patient's goals were averaged within occupational groups for both month one and month five. No pre-post therapy change was found with a Wilcoxon Signed Ranks test for

any group's ratings except those of the off-ward staff; off-ward staff rated significant improvement in the group of 22 patients. In view of the low inter-rater reliabilities and failure of other groups to rate the patients as improved, the significant improvement rated by the off-ward staff may best be attributed to chance fluctuations.

An examination of the month one ratings of the patient's average goals revealed that the patients rated themselves more favorably than other patients, ward staff, or off-ward staff did. After the program, the patients rated themselves higher than other patients or ward staff rated them. Off-ward staff gave significantly higher ratings than patients or ward staff. The unit director gave higher ratings than patients or ward staff.

The failure of the raters to rate the patients low on the dimensions on which they were expected to improve may be related to the fact that most of the raters were blind with respect to which items applied to which patients. Most studies of goal attainment are not conducted under blind conditions. The high initial ratings could mean that the patients don't have the problems that correspond to the goals. A further possibility is that the patients actually changed early in the first month and that the rating of the entire month obscured this change; of course normal clinical evaluation of patients cover much longer periods.

Nevertheless, as the goals employed in the study were those that the STU treatment staff commonly chose for patients, albeit in a less formal manner, the results of this study imply that the terms commonly used to describe patient change are either context-specific or ambiguous and, therefore, unsuited for research into patient dangerousness.

Behavior modification programs offer better chances to obtain objective measures of patient progress in maximum security institutions because of the daily observation of simple behaviors which they entail. The Activity Treatment Unit (ATU) of Oak Ridge has maintained such programs for over six years and has generated a large amount of useable data. In our first ATU study (Quinsey & Sarbit, 1975), we found that points earned daily for room care, self care, and ward work and weekly for mood and cooperation ratings were sufficiently sensitive to detect improvement among 12 chronic patients associated with a change in the program such that points were calculated daily rather than weekly. Using similar measures, Quinsey, Rice and Houghton (1978) followed the progress of 130 newly admitted men for 12 weeks of treatment in a ward token economy. Patients who were high point earners in the initial two weeks tended to be high point earners in the final six weeks. Among those patients who were low point earners in the first two weeks, those who were married, had charges leading to admission (as opposed to being transferred from another psychiatric hospital), were paranoid schizophrenic or otherwise psychotic, and had an occupation were more likely to improve. There were very high intercorrelations among the ratings of patient mood, cooperation, ward work and room and self care scores which suggested that patients were being assessed on a unitary dimension of "psychiatric disturbance." The results indicated not only that the token program should be individualized to make it relevant for patients who are high point earners at the outset and to offer contingencies for patient's individual problems but also that these on-ward measures were not

suitable for research on dangerousness because of their lack of independence and specificity.

Our failure to find satisfactory measures of on-ward behavior which might be related to patient dangerousness led us to a more direct attack on the problem by attempting to record the assault frequency of ATU patients (Quinsey, 1977c; Quinsey & Varney, 1978). We initially used records of assaults which were kept in on-ward nursing notes. In a retrospective study of the hospital records of four highly assaultive patients, we found that the descriptions of the events which preceded the assaults were often incomplete and that the retrospective nature of our evaluation made it difficult to verify that an exhaustive sample of assaults had been obtained.

Our difficulties in obtaining satisfactory data on the most significant clinical problems for many ATU patients led us to design a research study of all the assaults within the unit. We monitored all the assaults occurring on the ATU for a one year period. The results of this study had a major impact on our thinking about assault frequency. The first finding was that assaults were much more restricted geographically than had been previously thought; 60% of the 198 assaults occurred on one ward and 90% occurred on the upper or more secure wards. More important, however, was the finding that 13% ($n = 18$) of the patients committed 61% of the assaults. These findings indicate that a treatment intervention designed to reduce assault frequency could be concentrated on a small number of patients on a single ward.

When we asked the aggressive patients and the staff member involved most closely with an assault why the event had occurred, we received widely discrepant reasons from the two sources. Patients cited patient teasing or staff provocations as the reasons for their assaults, whereas the staff most typically advanced "no apparent reason" as an explanation. Although the discrepancies between the explanations offered by the two groups can be accounted for in part by the patients attempting to shift the blame from themselves to others and because the patients were, of course, in a better position to observe their own motivation for a particular assault, the staff often seemed completely unaware of events which may have triggered assaults even when these events largely involved their own behavior. Both sets of respondents agreed, however, that social stimuli such as patient teasing or sanctions by ward staff for patient misbehavior were the major causes of assaults. Data, when available from other witnesses, confirmed the importance of frustrating social stimuli such as patient teasing or staff sanctions.

Recent discussions of dangerousness (e.g., Quinsey, Ambtman, & Pruesse, 1977) have emphasized the role of situational variables in assaultive acts. As a research strategy in the assessment of dangerousness, therefore, a demonstration that some situational or behavioral intervention affects assault frequency can provide strong evidence that this situational variable should be taken into account in future assessment. We have attempted to reduce assault frequency using several methods. Our first intervention involved a social skills training program for highly assaultive patients (Quinsey, 1977c; Quinsey & Varney, 1977, 1978); this work has been described recently elsewhere (Rice & Quinsey, 1978) and so will not be described here. In this social skill training program, it gradually became apparent that significant reductions in assault frequency

throughout the unit would depend not only on the successful modification of patient behavior but also on the alteration of the social system of which patient assaultiveness is a part. A number of research findings led us to this conclusion. The first was the observation that staff were much more likely to be victims of assaults than the patients, who were both more numerous and in closer physical proximity to other patients than the staff. Secondly, a staff person could very seldom point to circumstances which might have led the patient to assault him. Quite often these circumstances were simply not recognizable or observable to the staff, but even on occasions where there appeared to be clear indications that something was amiss, these indications would often not be noticed. Informally, we observed marked individual differences among the attendants in their sensitivities to warning signals emitted by patients. In addition, although we had no hard data, it was apparent that on a given ward, some staff were more likely to be victims than others.

If we conceptualize assaults as resulting from dyadic social interactions, rather than as phenomena emanating solely from the patients pathologies, then it makes sense to attempt to modify staff behavior since the majority of victims were staff and the staff, being neither psychotic nor retarded, ought to be more easily modifiable than the patients. It is important to understand that by emphasizing the modification of staff behaviors we are not suggesting that staff provoke assaults, but rather that some assaults could be avoided by staff learning to make appropriate responses to warning signals emitted by patients or by learning to interact with patients in a manner which minimizes the likelihood of an assault.

In an attempt to increase staff awareness of potentially assaultive interactions and modify their responses to these situations, we designed an "assault prevention training task force." The task force was designed so that it would: (a) encourage staff to examine in detail the events preceding assaults for clues as to how they could have been avoided; (b) take advantage of more experienced attendants' skills in avoiding altercations and (c) provide an opportunity for staff who are repeated victims to receive advice from attendants who are not frequently assault victims.

The task force involves a peer review of each physical altercation between an attendant or other staff and a patient. Each time a staff is assaulted by a patient, he is interviewed by a group of attendants, a psychologist and managerial (nursing and attendant series) staff. One attendant staff was chosen to represent each of the four unit wards on the basis of their ability to command respect from other staff. The managerial staff were chosen to represent all levels of the chain of command within Oak Ridge.

A comparison of assault frequency in the 800 days before the task force began with the 750 day post-task force period indicated no change in overall assault frequency. The average number of assaults per day was .496 in the pre-task force period and was .477 in the post-task force period. There was, however, a shift in the type of assault victim. Attendants were more likely than patients to be victims in the pre-task force period and less likely afterward (Chi square = 15.35, $df = 1$, $p < .001$). Although it is tempting to conclude on the basis of this result that the task force was responsible for the predicted decline in attendant victims, other explanations cannot be ruled out completely on the basis of

our data. In particular, the reason for the increase in patient-patient assaults is not immediately apparent. Further studies of more extensive staff training efforts in the areas of crisis intervention and effective restraint technique should allow us to evaluate the effects of our interventions more clearly.

Psychophysiological Assessments of Sexual Offenders

Perhaps the greatest progress in the assessment of institutionalized men has been made in the area of the measurement of inappropriate sexual preferences (Quinsey, 1973, 1977b). In agreement with studies done by other research teams, we have demonstrated that penile responses to slides of persons varying in age and sex can discriminate child molesters from normals and relate very closely to the child molesters' history of victim choice, whereas their verbal responses do not (Quinsey, Steinman, Bergersen, & Holmes, 1975). More recently we have found that non-incestuous child molesters have more inappropriate sexual age preferences than incestuous child molesters (Quinsey, Chaplin, & Carrigan, 1979). In a treatment study, we have shown that penile responses to child and adult categories change as a result of a classical conditioning form of aversion therapy (Quinsey, Bergersen, & Steinman, 1976). Unfortunately, all penile response measures of sexual preference have to be interpreted with caution as a substantial proportion of non-sex offenders can increase their penile responses to children and/or decrease their responses to adult women in accord with instructions (Quinsey & Bergersen, 1976), even when the slides are accompanied by auditory descriptions of relevant sexual fantasies (Quinsey & Carrigan, 1978).

The extent of the "faking" problem with child molesters is moot as we routinely find psychophysiological evidence of inappropriate sexual preferences among child molesters who claim to prefer adults as sexual partners (Quinsey, Steinman, Bergersen, & Holmes, 1975). Regardless of this consideration, the utility of penile response measures of sexual preference in assessment can best be evaluated by relating these responses (and changes in these responses which are correlated with some therapeutic intervention) to post-release data.

The validation of sexual preference profiles and changes in these profiles through follow-up studies presents formidable methodological difficulties. First, one must demonstrate changes in sexual preference which are statistically significant within individual patients, otherwise the "change" scores are uninterpretable and one would logically predict recidivism. The amount of change which results from a treatment appears to be critically dependent on the details of the procedure. In our first study we failed to obtain statistically significant individual shifts in most of our patients treated with a classical conditioning aversion therapy procedure (Quinsey, Bergersen, & Steinman, 1976); however, using a procedure in which a subject receives feedback regarding his penile responses and subsequently (if the biofeedback procedure fails) electric shock contingent upon penile responses to child slides, we have obtained significant improvements in pre-post generalization probes in more than half of the child molesters that we treated (Quinsey, Chaplin, & Carrigan, 1978).

A further problem in the validation of such measures is the low base rate of

new offenses against children among child molesters released from maximum security psychiatric institutions. In a follow-up currently in progress of 119 released child molesters who had been assessed and/or treated, it has been found that only seven were convicted of, returned to Oak Ridge because of, or were known by the staff of regional mental health centres to have committed, a new sexual offense against children.

Of course it may be argued that the relationship between inappropriate sexual age preferences and the commission of new sex offenses against children should not be straightforward anyway because other factors, such as poor heterosocial skills and preferences for deviant acts, are also involved. We have attempted to deal with both of these issues in our research; our efforts in the area of social competence have been reviewed elsewhere recently (Rice & Quinsey, 1978) and, therefore, will not be discussed in this context.

The importance of assessing and modifying sexual arousal to inappropriate acts is dramatically illustrated by a patient who was referred for treatment because he had sexually assaulted and attempted to mutilate a young boy with a knife. Psychophysiological assessment of the patient showed that sexual arousal was elicited by slides of pubescent and child males. Biofeedback training, in which the patient's penile responses were fed back to him via lights underneath a rearview projection screen helped him to acquire some control of his penile responses — i.e., to become more aroused to slides of adult males and females (the patient wished to become bisexual), and less aroused to slides of young boys. Nevertheless, after this treatment the patient still reported frequent masturbation to sadistic fantasies of mutilating and/or killing older boys and young men. Thus only his sexual age preference had been affected. Using a satiation procedure similar to that of Marshall and Lippens (1977), in which the patient masturbated while verbalizing his deviant fantasies for 90 minute periods regardless of whether he had reached orgasm, we were able to affect substantial reductions in his sexual arousal to audio-taped descriptions of his deviant fantasies while maintaining his arousal to non-sadistic fantasy material (Quinsey & Chaplin, 1978). This case study clearly indicates the need to assess and modify the preferences that sexual offenders may have for inappropriate acts as well as inappropriately aged sexual partners.

Rapists clearly fall into the category of persons who perform inappropriate acts with appropriately aged partners. Using penile responses to audiotaped material, Abel, Barlow, Blanchard and Guild (1977) found that rapists showed greater sexual arousal to rape scenes than consenting sex scenes whereas this was not true of men with other sexual difficulties. Using a similar paradigm, Barbaree, Marshall and Lanthier (1978) demonstrated that male graduate students could be discriminated from rapists on the basis of penile responses to rape scenes but not using responses to consenting sex scenes. We have replicated the essential findings of these two studies by showing that Oak Ridge rapists can be differentiated from a control group comprised of non-sex offender patients and community volunteers of working class backgrounds (Quinsey, Chaplin, & Varney, 1979). The use of audiotaped stimuli in the psychophysiological measurement of the sexual arousal of rapists, as well as other sexual offenders, makes possible easy individualization of treatment and assessment and allows

the measurement of arousal generated by an extremely wide variety of sexual cues. As deviant sexual fantasies appear to play a central role in sex offenses, the importance of such a measurement technique cannot be underestimated.

Conclusions

Certainly no one would accuse us of having solved the problems involved in predicting dangerousness; nor could we claim even to have solved the assessment problem in the limited areas where we have concentrated our research. Nevertheless, if one defines progress in science as the accumulation of replicable research findings which are orderly in the sense of being related to each other in a coherent manner, then I think that progress, however undramatic, has been made in our research and that of others who are doing similar work. Certainly, this review of our research projects indicates that progress has been very uneven in the various areas which we have investigated. The amount of progress in these various areas seems to follow a definite pattern in that the magnitude of the effects of the various independent variables and the orderliness of the data seem to be a direct function of the extent to which the dependent measures are quantifiable and theoretically relevant to the antisocial behavior under investigation, as well as the extent to which the offender or subject population has been homogenized through selection on theoretically relevant variables.

Several examples will serve to illustrate the point. Prediction of dangerousness studies using demographic data have not progressed because of the heterogeneity of the population studied, the coarseness of the predictive variables and the "noise" inherent in the measurement of recidivism. Research on the psychophysiological assessment of sex offenders' sexual preferences, however, has made progress because the measure of sexual arousal is theoretically relevant and quantifiable and the population studies can be divided rationally according to the type of deviant behavior and deviant object choice. In fact, these sorts of data become most intelligible when analyzed at the level of individual arousal patterns.

It is my view, based upon the experience summarized above, that research based on "criminals" or "mentally disordered offenders" using independent variables which have no compelling theoretical relevance and dependent measures which are imprecise is doomed to certain failure, although in all likelihood this type of research will continue for some time to come.

References

Abel, G. G., Barlow, D. H., Blanchard, E. G., & Guild, D. The components of rapist's sexual arousal. *Archives of General Psychiatry*, 1977, 34, 895-903.

Arnold, L., Quinsey, V. L., & Velner, I. Overcontrolled hostility among men found not guilty by reason of insanity. *Canadian Journal of Behavioral Science*, 1977, 9, 333-340.

Barbaree, H. E., Marshall, W. L., & Lanthier, R. D. *Deviant sexual arousal in rapists*. Unpublished manuscript, 1978.

Barker, E. T., & Buck, M. F. LSD in a coercive milieu therapy program. *Canadian Psychiatric Association Journal*, 1977, 22, 311-314.

Barker, E. T., & Mason, M. H. Buber behind bars. *Canadian Psychiatric Association Journal*, 1968, 13, 61-72. (a)

Barker, E. T., & Mason, M. H. The insane criminal as therapist. *Canadian Journal of Corrections*, 1968, 10, 553-561. (b)

Barker, E. T., Mason, M. H., & Wilson, J. Defence-disrupting therapy. *Canadian Psychiatric Association Journal*, 1969, 14, 355-359.

Barker, E. T., & McClaughlin, A. J. The total encounter capsule. *Canadian Psychiatric Association Journal*, 1977, 22, 355-360.

Gulevich, G. D., & Bourne, P. G. Mental illness and violence. In D. N. Daniels, M. F. Gilula, & F. M. Ochburg (Eds.), *Violence and the struggle for existence*. Boston: Little, Brown, 1970.

Marshall, W. L., & Lippens, K. The clinical value of boredom: A procedure for reducing inappropriate sexual interests. *Journal of Nervous and Mental Disease*, 1977, 165, 283-287.

Mergargee, E. J. The prediction of dangerous behavior. *Criminal Justice and Behavior*, 1976, 3, 3-21.

Mergargee, E. J., Cook, P. E. & Mendelsohn, G. A. Development and evaluation of an MMPI scale of assaultiveness in overcontrolled individuals. *Journal of Abnormal Psychology*, 1967, 72, 519-528.

Mesnikoff, A., & Lauterbach, C. G. The association of violent dangerous behavior with psychiatric disorders: A review of the research literature. *Journal of Psychiatry and Law*, 1975, 3, 415-445.

Monahan, J. The prediction of violence. In D. Chappell and J. Monahan (Eds.), *Violence and criminal justice*. Lexington, Mass.: Lexington Books, 1975.

Pruesse, M., & Quinsey, V. L. The dangerousness of patients released from maximum security: A replication. *Journal of Psychiatry and Law*, 1977, 5, 293-299.

Quinsey, V. L. *A note on recording psychiatric patients' progress*. Unpublished manuscript, 1972.

Quinsey, V. L. Methodological issues in evaluating the effectiveness of aversion therapies for institutionalized child molesters. *Canadian Psychologist*, 1973, 14, 350-361.

Quinsey, V. L. Psychiatric staff conferences of dangerous mentally disordered offenders. *Canadian Journal of Behavioural Science*, 1975, 7, 60-69.

Quinsey, V. L. Problems in the treatment of mentally disordered offenders. *Canada's Mental Health*, 1977, 25, 2-3. (a)

Quinsey, V. L. The assessment and treatment of child molesters: A review. *Canadian Psychological Review*, 1977, 18, 204-220. (b)

Quinsey, V. L. Studies in the reduction of assaults in a maximum security psychiatric institution. *Canada's Mental Health*, 1977, 25, 21-23. (c)

Quinsey, V. L., & Ambtman, R. Psychiatric assessments of the dangerousness of mentally ill offenders. *Crime and Justice*, 1979. (a)

Quinsey, V. L., & Ambtman, R. Variables affecting psychiatrists' and teachers' assessments of the dangerousness of mentally ill offenders. *Journal of Consulting and Clinical Psychology*, 1979. (b)

Quinsey, V. L., Ambtman, R., & Pruesse, M. *Institutional release policy and the identification of dangerous men*. To appear in a proceedings volume of the Symposium on Violence in Canadian Society, Qualicum Beach, Vancouver Island, March, 1977.

Quinsey, V. L., & Arnold, A. *MMPI profiles of men referred for a pre-trial assessment as a function of offence type*. Unpublished manuscript, 1978.

Quinsey, V. L., & Bergersen, S. G. Instructional control of penile circumference. *Behavior Therapy*, 1976, 7, 489-493.

Quinsey, V. L., Bergersen, S. G., & Steinman, C. M. Changes in physiological and verbal responses of child molesters during aversion therapy. *Canadian Journal of Behavioural Science*, 1976, 8, 202-212.

Quinsey, V. L., & Boyd, B. A. An assessment of the characteristics of patients held on Warrants of the Lieutenant Governor. *Crime and Justice*, 1977, 4, 268-274.

Quinsey, V. L., & Carrigan, W. F. Penile responses to visual stimuli: Instructional control with and without auditory sexual fantasy correlates. *Criminal Justice and Behavior*, 1978, 5, 212-220.

Quinsey, V. L., & Chaplin, T. C. *Satiation as treatment for sadistic fantasies: A case study*. Unpublished manuscript, 1978.

Quinsey, V. L., Chaplin, T. C. & Carrigan, W. F., *Biofeedback and signaled punishment in the modification of inappropriate sexual age preferences*. Presented at the meeting of the Ontario Psychological Association, Ottawa, February, 1978.

Quinsey, V. L., Chaplin, T. C., & Carrigan, W. F. Sexual preferences among incestuous and non-incestuous child molesters. *Behavior Therapy*, 1979, in press.

Quinsey, V. L., Chaplin, T. C. & Varney, G. W. A comparison of rapists' and non-sex offenders' sexual preferences for mutually consenting sex, rape, and sadistic acts. Presented at the Conference on Sexual Aggressives, New York City, May, 1979.

Quinsey, V. L., & Harris, G. T. *Goal attainment ratings of patient change in a maximum security social therapy program*. Unpublished manuscript, 1976.

Quinsey, V. L., Pruess, M., & Fernley, R. Oak Ridge patients: Prerelease characteristics and postrelease adjustment. *Journal of Psychiatry and Law*, 1975, 3, 63-77. (b)

Quinsey, V. L., Ricc, M. E., & Houghton, R. *Predicting treatment outcome in a maximum security token economy*. Presented at the meeting of the Canadian Psychological Association in Ottawa, June, 1978.

Quinsey, V. L., & Sarbit, B. Behavioral changes associated with the introduction of a token economy in a maximum security psychiatric institution. *Canadian Journal of Criminology and Corrections*, 1975, 17, 177-182.

Quinsey, V. L., Steinman, C. M., Bergersen, S. G., & Holmes, T. F. Penile circumference, skin conductance, and ranking responses of child molesters and "normals" to sexual and nonsexual visual stimuli. *Behavior Therapy*, 1975, 6, 213-219.

Quinsey, V. L., & Varney, G. W. Social skills game: A general method for the modeling and practice of adaptive behaviors. *Behavior Therapy*, 1977, 8, 279-281.

Quinsey, V. L., & Varney, G. W. Characteristics of assaults and assaulters in a maximum security psychiatric unit. *Crime and Justice*, 1978, 5, 212-220.

Quinsey, V. L., Varney, G. W., & McCann, D. *A laboratory comparison of different methods of response reduction among assaultive and nonassaultive psychiatric patients*. Presented at the meeting of the Ontario Psychological Association in Ottawa, February, 1978.

Quinsey, V. L., Warneford, A., Pruess, M., & Link, N. Released Oak Ridge patients: A follow-up study of review board discharges. *British Journal of Criminology*, 1975, 15, 264-270.

Rice, M. E., & Quinsey, V. L. *The training of social competence in dangerous male psychiatric patients*. Unpublished manuscript, 1978.

Shah, S. A. Dangerousness: a paradigm for exploring some issues in law and psychology. *American Psychologist*, 1978, 33, 224-238.

Steadman, H. J. Predicting "dangerousness." In D. J. Madden and J. R. Lion (Eds.), *Rage, hate, assault and other forms of violence*. New York: Spectrum, 1976.

Steadman, H. J., & Cocozza, J. J. *Careers of the criminally insane*. Lexington, Mass.: Lexington Books, 1974.

Tong, J. E., & McKay, J. W. A statistical follow-up of mental defectives of dangerous or violent propensities. *British Journal of Delinquency*, 1959, 9, 276-284.

Waller, J. *Men released from prison*. Toronto: University of Toronto Press, 1974.

Wenk, E., Robison, J. O., & Smith, G. W. Can violence be predicted? *Crime and Delinquency*, 1972, 18, 393-402.

Wiggins, J. S. *Personality and prediction: Principles of personality assessment*. Don Mills, Ontario: Addison-Wesley, 1973.